

PROJETO EM CIÊNCIA DE DADOS

SUMÁRIO

SEMESTRE	2024/2
PROJETO	Análise dos mesários das eleições de 2020 e 2024 – Porto Alegre e Belém
COMPONENTES DO GRUPO	Ana Cristina Follmann Schmidt Jeniffer Moreira Borges

Breve descrição do problema

Identificar as características dos mesários nas eleições de prefeitos e vereadores de 2024 nas cidades de Porto Alegre e Belém.

Breve descrição da solução proposta

Identificar as bases de arquivos referentes aos mesários do Pará e do Rio Grande do Sul. Após isso, filtrar Belém e Porto Alegre. Remover as colunas que não serão utilizadas na análise, verificar valores faltantes e valores estatísticos. Fazer um dashboard e relatório com as análises.

Fases da Metodologia CRISP-DM

Fase	Tarefas	% Conclusão
1. Entendimento do Negócio	<ul style="list-style-type: none">- Definição dos objetivos do projeto- Compreensão das necessidades de visualização e métricas principais para o dashboard- Formulação de perguntas e hipóteses iniciais	100%
2. Compreensão dos Dados	<ul style="list-style-type: none">- Coleta dos dados disponíveis- Identificação e descrição das fontes de dados- Exploração inicial dos dados (análise estatística e de qualidade)- Avaliação da completude e consistência	100%

3. Preparação dos Dados	<ul style="list-style-type: none">- Limpeza dos dados (remoção de valores ausentes e inconsistências)- Transformação de colunas para melhorar a visualização- Integração das bases de dados para análise consolidada- Seleção das variáveis para visualização no Power BI	90%
4. Modelagem	<ul style="list-style-type: none">- Definição de métricas e KPIs a serem calculados no Power BI- Seleção de gráficos e layouts adequados para as necessidades do negócio- Construção de cálculos e agregações necessários para as visualizações	0%
5. Avaliação	<ul style="list-style-type: none">- Comparação das visualizações com os objetivos iniciais- Validação da clareza e eficiência do dashboard	0%
6. Implantação	<ul style="list-style-type: none">- Criação do dashboard interativo no Power BI- Documentação dos relatórios e dos filtros disponíveis	0%

Resumo do que foi concluído até o momento

Até o momento, o projeto avançou significativamente nas fases iniciais do CRISP-DM, focando principalmente no Entendimento do Negócio, Compreensão dos Dados e Preparação dos Dados. O próximo grande desafio será a transição para o Power BI, onde o foco estará em traduzir os dados em visualizações intuitivas e interativas.

Autocrítica

Aderência ao CRISP-DM, com execução organizada.

Nota do grupo: 9.0

Motivo: Poderia melhorar na organização.

RELATÓRIO

1. Compreensão dos Dados

Coleta dos dados

Os dados foram obtidos do repositório de dados abertos do TSE, nos datasets de mesários dos anos de 2020 e 2024.

<https://dadosabertos.tse.jus.br/dataset/mesarios-2024> e

<https://dadosabertos.tse.jus.br/dataset/mesarios-2020>

Descrição dos dados

Coleta inicial:

O dataset original contém 19 colunas com informações sobre os mesários convocados para atuar nas eleições. Abaixo estão as descrições de cada coluna:

DT_Geracao: Data de geração do arquivo, indicando o dia em que o dataset foi criado.

HH_Geracao: Hora de geração do arquivo, indicando a hora em que o dataset foi criado.

AA_Eleicao: Ano da eleição em que os mesários foram convocados.

Nr_Turno: Número do turno da eleição (1 para o primeiro turno, 2 para o segundo turno, etc.).

SG_UF: Sigla da Unidade Federativa (UF) onde o mesário foi convocado para atuar.

CD_Municipio: Código do município, um identificador único para o município de atuação.

NM_Municipio: Nome do município onde o mesário foi convocado para atuar.

NR_Zona: Número da zona eleitoral em que o mesário atuou.

CD_Genero: Código referente ao gênero do mesário.

DS_Genero: Descrição do gênero do mesário (exemplo: "Masculino", "Feminino").

CD_Estado_Civil: Código referente ao estado civil do mesário.

DS_Estado_Civil: Descrição do estado civil do mesário (exemplo: "Solteiro", "Casado").

CD_Faixa_Etaria: Código que identifica a faixa etária do mesário.

DS_Faixa_Etaria: Descrição da faixa etária (exemplo: "21 a 24 anos", "40 a 44 anos").

CD_Grau_Escolaridade: Código que representa o nível de escolaridade do mesário.

DS_Grau_Escolaridade: Descrição do nível de escolaridade (exemplo: "Ensino Médio Completo", "Superior Completo").

TP_Convocaçao: Tipo de convocação do mesário, indicando se ele foi voluntário ou não (exemplo: "Voluntário", "Não voluntário").

TP_Comparecimento: Tipo de comparecimento do mesário, indicando se ele compareceu ou não no dia da eleição (exemplo: "Comparecimento", "Falta").

QT_Mesarios_Perfil: Quantidade de mesários com o mesmo perfil de convocação e características demográficas.

Mesários Belém 2020: 5.680 registros

Mesários Belém 2024: 5.683 registros

Mesários Porto Alegre 2020: 5.639 registros

Mesários Porto Alegre 2024: 6.169 registros

Análise exploratória dos dados

Dados estatísticos de forma resumida:

Belém 2020:

- **Total de Registros:** 5.680.
- **Distribuição de Gênero:** Predominância de mesários de gênero "Feminino" representando 3.120 registros.
- **Estado Civil Mais Comum:** "Solteiro", com 3.202 registros.
- **Faixa Etária Mais Frequente:** "35 a 39 anos" (764 registros).
- **Escolaridade Predominante:** "Superior Completo" (1.843 registros).
- **Tipo de Convocação:** A maioria dos mesários foram "Não voluntários" (3.031 registros).
- **Tipo de Comparecimento:** A maioria compareceu, com 4.552 registros marcados como "Comparecimento".

Belém 2024:

- **Total de Registros:** 5.683.
- **Distribuição de Gênero:** A maioria é feminina, com 3.299 registros.

- **Estado Civil Mais Comum:** "Solteiro", com 3.146 registros.
- **Faixa Etária Mais Frequente:** "45 a 49 anos" (691 registros).
- **Escolaridade Predominante:** "Ensino Médio Completo" (1.740 registros).
- **Tipo de Convocação:** Mais da metade dos mesários foram "Voluntários" (3.091 registros).
- **Tipo de Comparecimento:** 5.233 registros indicaram "Comparecimento".

Porto Alegre 2020:

- **Total de Registros:** 5.639.
- **Distribuição de Gênero:** Mais mulheres, com 3.086 registros.
- **Estado Civil Mais Comum:** "Solteiro", com 3.256 registros.
- **Faixa Etária Mais Frequente:** "40 a 44 anos" (821 registros).
- **Escolaridade Predominante:** "Superior Completo" (1.565 registros).
- **Tipo de Convocação:** Predomínio de "Não voluntários" (3.188 registros).
- **Tipo de Comparecimento:** A maioria compareceu, com 5.298 registros.

Porto Alegre 2024:

- **Total de Registros:** 6.169.
- **Distribuição de Gênero:** Predominância de mulheres, com 3.490 registros.
- **Estado Civil Mais Comum:** "Solteiro", com 3.502 registros.
- **Faixa Etária Mais Frequente:** "45 a 49 anos" (840 registros).
- **Escolaridade Predominante:** "Superior Completo" (1.756 registros).
- **Tipo de Convocação:** Maioria "Não voluntários" (3.100 registros).
- **Tipo de Comparecimento:** A maior parte dos mesários compareceu (6.113 registros).

Verificação de qualidade dos dados

Dados adequados para a análise proposta.

2. Preparação dos Dados

Limpeza dos dados

Os datasets não contêm dados nulos. Algumas colunas foram removidas: DT_Geracao, HH_Geracao, CD_Municipio, NR_Zona, CD_Genero, CD_Estado_Civil, CD_Faixa_Etaria, CD_Grau_Escolaridade.

As colunas removidas não serão necessárias no dashboard a ser criado. As bases dos mesários de 2020 e 2024 foi combinada em um único arquivo, formando uma única base.

Criação de atributos e registros

Não ocorreu a criação de novas instâncias.

Integração de dados

Foram utilizadas no total quatro bases de dados oficiais do TSE, duas de 2020 e duas de 2024, referentes aos estados do Pará e do Rio Grande do Sul. As bases foram filtradas, para serem localizadas as cidades de Belém e Porto Alegre. Após isso, o tratamento foi realizado, e novas quatro bases foram firmadas: Mesários - Belém 2020, Mesários - Belém 2024, Mesários - Porto Alegre 2020 e Mesários - Porto Alegre 2024. A união dessas quatro bases formou o dataset final.

Descrição do dataset final

O dataset final contém os registros das cidades de Porto Alegre e Belém, referentes aos anos de 2020 e 2024, devidamente limpos e pré-processados.

Colunas Finais: Ano da eleição, turno, Unidade Federativa (UF), município, zona eleitoral, gênero, estado civil, faixa etária, grau de escolaridade, tipo de convocação, tipo de comparecimento e quantidade de mesários.

O dataset está agora pronto para a fase de modelagem e criação de visualizações no Power BI, com dados consistentes e normalizados.

3. Autocrítica

A nota atribuída pelo grupo é 9,0. Consideramos que o desempenho até o momento foi bom, com um progresso organizado e bem alinhado com os objetivos do projeto. No entanto, identificamos que a organização interna pode ser aprimorada.

O grupo está confiante de que conseguirá concluir o projeto de maneira satisfatória e alcançar a entrega de 100% das etapas planejadas.